

Fig.1

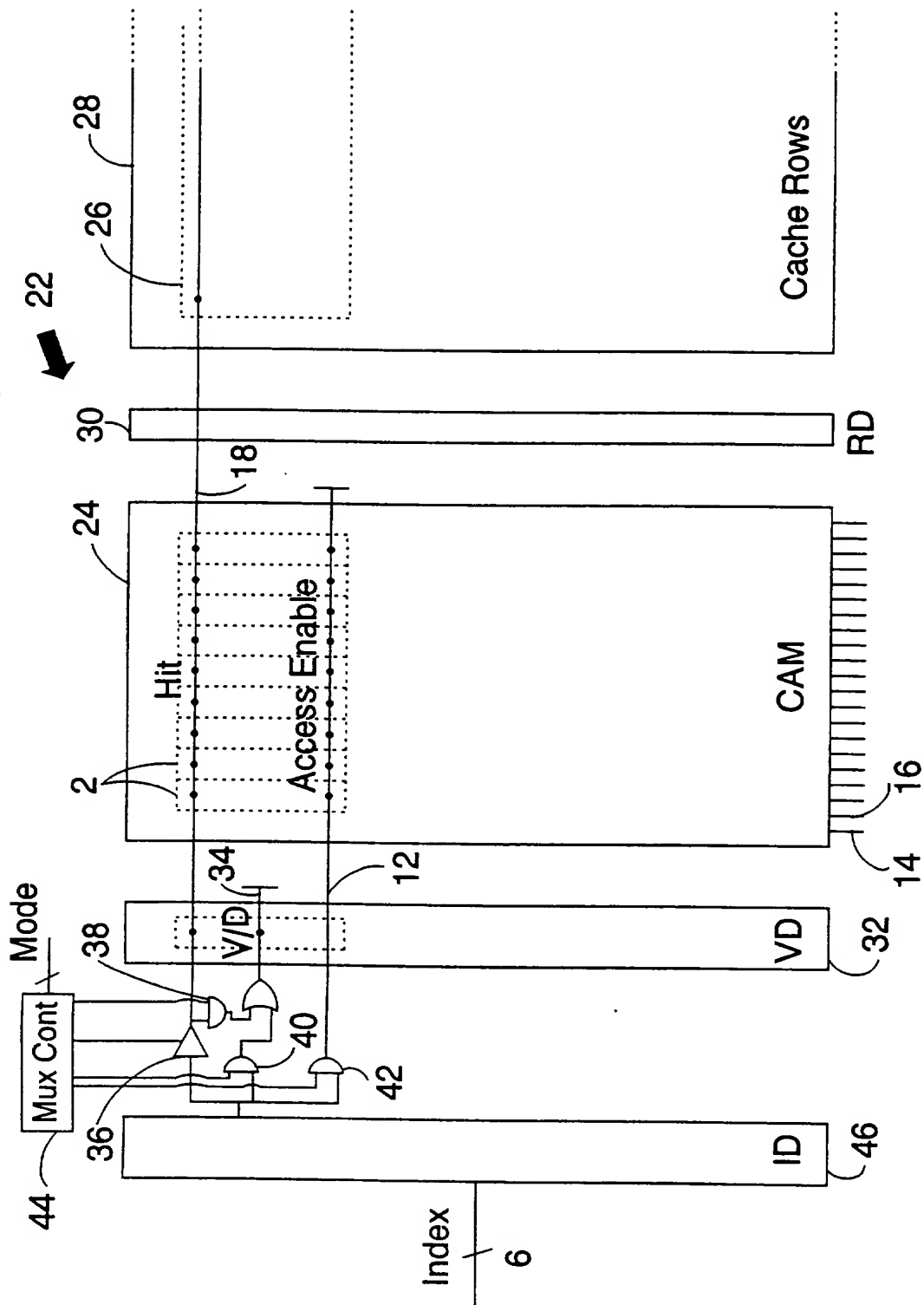


Fig. 2

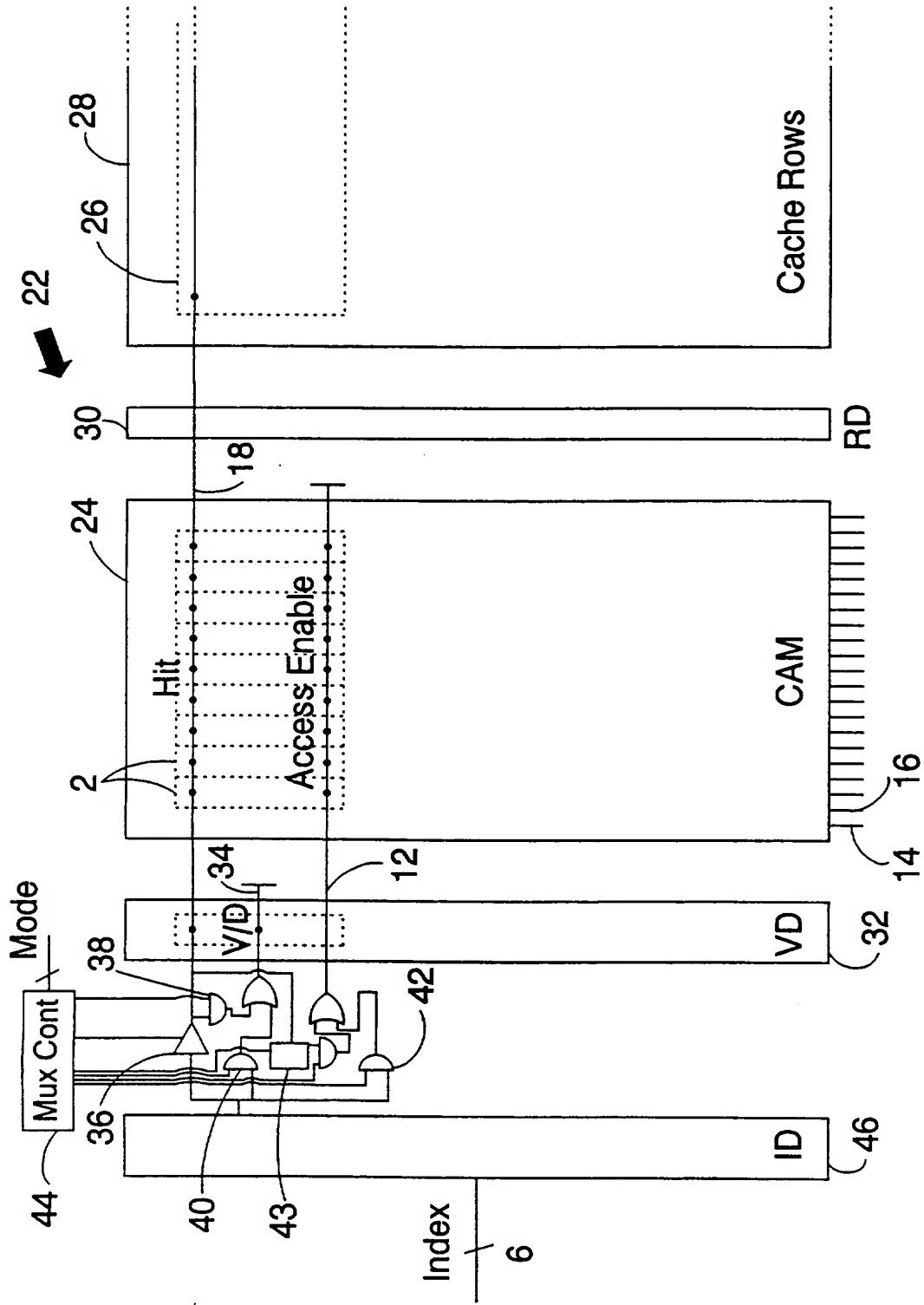


Fig.3

5

CACHE MEMORY

This invention relates to the field of data processing. More particularly, this invention relates to data processing systems incorporating a cache memory. Cache memories incorporating content addressable memories (CAMs) and associated RAM
10 memories are known within data processing systems. A CAM may be used to provide storage for a TAG within a cache memory system. A CAM allows an input value to be compared in parallel with a plurality of values stored in respective content addressable storage rows within the CAM. If a content addressable storage row is storing a content addressable data word matching the input data value, then a hit
15 signal is generated for that row signalling a match. The hit signal is then used to select a corresponding cache row within the RAM of the cache memory that is storing data values having memory addresses indicated by the TAG value.

Another part of a cache memory is the victim select circuitry that controls
20 which cache data is to be removed from the cache memory when new cache data needs to be stored into the cache memory and the cache memory is already full. Various victim selection strategies are known, such as random, least recently used and round-robin.

25 A constant aim within data processing systems is to reduce the size of the circuits. This increases speed, reduces cost and reduces power consumption as well as providing other advantages.

Viewed from one aspect the present invention provides apparatus for
30 processing data having a cache memory, said cache memory comprising:

a plurality of content addressable storage rows, each content addressable storage row containing a plurality of content addressable bit storage cells and serving to store a content addressable data word;

a plurality of bit lines connecting corresponding content addressable bit
35 storage cells in different content addressable storage rows such that an input data word

5 on said plurality of bit lines is compared in parallel with the contents of each connected content addressable storage row;

a plurality of hit lines, each content addressable storage row having an associated hit line upon which a hit signal is asserted if said input data word matches a content addressable data word stored within said content addressable storage row;

10 a plurality of cache data rows each storing a row of cache data, said content addressable memory serving to store address data to identify memory addresses for corresponding rows of cache data such that a hit signal generated on a hit line in said content addressable memory enables access to a corresponding cache data row within said cache memory; and

15 an index decoder for decoding an index value to generate a select signal for selecting a content addressable storage row and a cache data row for replacement within said cache memory; wherein

said select signal is passed to said cache data row via a corresponding one of said hit lines.

20

It is known within cache memory systems to provide signal lines running across each content addressable storage row. These may be a hit line for indicating a match with a TAG value, a write enable for enabling writing to a content addressable storage row and a corresponding cache data row, and a read enable to enable reading from a content addressable storage row. Either the hit signal or the read enable signal needs to be passed to a cache data row and so a multiplexer is typically provided between the CAM and the cache RAM to select one of these signals to be passed to the RAM. Such a multiplexer may be on the critical path limiting the maximum speed of operation of the cache memory and accordingly its presence is disadvantageous. A further disadvantage of this known arrangement is that the three separate signal lines passing across the content addressable storage row limit the minimum size that can be achieved for the content addressable storage row.

The present invention recognises both of the above problems and realises that they may be overcome by using the hit line traversing the content addressable storage row to also pass a victim select signal to a victim cache data row. This can reduce the

5 number of signal lines passing across each content addressable storage row from three to two and remove a multiplexer from the critical path.

In preferred embodiments of the invention access enable lines are also provided across each content addressable storage row to provide access to the contents
10 of the content addressable memory when it is desired that this should be independent of any signal upon the hit line.

The content addressable memory may also have valid and dirty bits associated with each row and in this circumstance it is desirable that a selectable connection is
15 provided for coupling a hit signal to enable access to any relevant valid and dirty bit in a matching row.

Viewed from another aspect the present invention provides a method of processing data using a cache memory, said comprising method comprising the steps
20 of:

storing content addressable data words within a plurality of content addressable storage rows, each content addressable storage row containing a plurality of content addressable bit storage cells;

connecting corresponding content addressable bit storage cells in different
25 content addressable storage rows with a plurality of bit lines such that an input data word on said plurality of bit lines is compared in parallel with the contents of each connected content addressable storage row;

providing a plurality of hit lines, each content addressable storage row having an associated hit line upon which a hit signal is asserted if said input data word
30 matches a content addressable data word stored within said content addressable storage row;

providing a plurality of cache data rows each storing a row of cache data, said content addressable memory serving to store address data to identify memory addresses for corresponding rows of cache data such that a hit signal generated on a
35 hit line in said content addressable memory enables access to a corresponding cache data row within said cache memory; and

5 providing an index decoder for decoding an index value to generate a select signal for selecting a content addressable storage row and a cache data row for replacement within said cache memory; wherein

said select signal is passed to said cache data row via a corresponding one of said hit lines.

10

The present invention may also be used to address the problem of increasing processing speed when a block transfer of data from one memory location to another memory location is required with that data then being used from the new location. Known systems have required the data to be read from the main memory to the processor, written back from the processor to the new address within the main
15 memory and finally read from the main memory or processor back to the cache memory with the associated TAG for the new memory location ready for further use. This is a relatively slow process.

20 Viewed from another aspect the present invention provides apparatus for processing data having a cache memory, said cache memory comprising:

a plurality of content addressable storage rows, each content addressable storage row containing a plurality of content addressable bit storage cells and serving to store a content addressable data word;

25 a plurality of bit lines connecting corresponding content addressable bit storage cells in different content addressable storage rows such that an input data word on said plurality of bit lines is compared in parallel with the contents of each connected content addressable storage row;

a plurality of hit lines, each content addressable storage row having an associated hit line upon which a hit signal is asserted if said input data word matches a
30 content addressable data word stored within said content addressable storage row;

a plurality of access enable lines, each content addressable storage row having an associated access enable line upon which an access signal may be asserted to allow a content addressable data word stored within said content addressable storage row to
35 be accessed;

5 a plurality of cache data rows each storing a row of cache data, said content addressable memory serving to store address data to identify memory addresses for corresponding rows of cache data such that a hit signal generated on a hit line in said content addressable memory enables access to a corresponding cache data row within said cache memory;

10 wherein, in an access mode, a connection is provided between a hit line for a content addressable storage row and an access enable line for that content addressable storage row such that an input data word on said bit lines can be compared with said content addressable data words within said plurality content addressable storage rows and if a match occurs for a content addressable storage row generating a hit signal to
15 enable access to that row by passing said hit signal to serve as an access enable signal upon an access enable line.

Providing a connection between the hit signal line and the access enable signal line allows the data stored within the content addressable memory to be manipulated
20 even when its location within the content addressable memory is not known, i.e. the content addressable memory can first be searched to find the data item in question to generate a hit signal and then this hit signal be redirected to provide access for manipulation to the relevant row within the content addressable memory. This feature allows the data stored within a content addressable memory to be manipulated in situ
25 in a manner not previously possible. One example of such a manipulation is to modify the TAG value associated with a row of cache data while the cache data is still stored within the cache memory RAM. More specifically, words from a first memory location can be read to the cache memory, the TAG value stored within the content addressable memory found and modified to point to a second memory location and
30 the cache words directly from the cache memory using their associated second memory address without having to first be written back to the main memory. The dirty bit associated with a cache memory row that is so modified can be set to ensure that when the cache data is removed from the cache memory it is written back to the main memory such that the integrity of the system is maintained.

5 Viewed from a further aspect the invention provides a method of processing data using a content addressable memory, said method comprising the steps of:

storing content addressable data words within a plurality of content addressable storage rows, each content addressable storage row containing a plurality of content addressable bit storage cells;

10 connecting corresponding content addressable bit storage cells in different content addressable storage rows with a plurality of bit lines such that an input data word on said plurality of bit lines is compared in parallel with the contents of each connected content addressable storage row;

providing a plurality of hit lines, each content addressable storage row having
15 an associated hit line upon which a hit signal is asserted if said input data word matches a content addressable data word stored within said content addressable storage row;

providing a plurality of access enable lines, each content addressable storage row having an associated access enable line upon which an access signal may be
20 asserted to allow a content addressable data word stored within said content addressable storage row to be accessed;

wherein, in an access mode, a connection is provided between a hit line for a content addressable storage row and an access enable line for that content addressable storage row such that an input data word on said bit lines can be compared with said
25 content addressable data words within said plurality content addressable storage rows and if a match occurs for a content addressable storage row generating a hit signal to enable access to that row by passing said hit signal to serve as an access enable signal upon an access enable line.

30 Embodiments of the invention will now be described, by way of example only, with reference to the accompanying drawings in which:

Figure 1 illustrates a content addressable memory cell;

35 Figure 2 illustrates a cache memory incorporating a content addressable memory according to a first embodiment; and

5

Figure 3 illustrates a cache memory incorporating a content addressable memory according to a second embodiment.

Figure 1 shows a content addressable memory cell 2. A bit value is stored by two invertors 4, 6. Gating transistors 8, 10 provide access control to the invertors 4, 6. When the gating transistors 8, 10 are turned on by an access enable signal RD/WR on the access enable signal line 12, then bit values on the bit lines 14, 16 (complementary values) are passed to or read from the invertors 4, 6 via transmission gates 5, 7.

During a comparison operation, an input data word is applied to the bit line pairs 14, 16 and compared in parallel within each of the content addressable memory rows with the data value stored by all the invertors 4, 6, making up that row. The hit lines 18 for each row are precharged high. If a memory cell 2 does not contain a matching bit to that input upon the bit line pairs 14, 16, then the transistor 20 is turned on so discharging the associated hit line 18. Accordingly, if any content addressable memory row is storing a value for which all of the bits match the input data word, then its hit line 18 will remain at the precharged high value and this then serves as a hit signal indicating the match. A single non-matching bit within a row results in a discharge of the hit line for that row. When the content addressable memory cell 2 is not being accessed for a read or write, the bit lines 14, 16 are both driven to zero. One of the transmission gates 5, 7 will then be enabled, but the transistor 20 will not be enabled and so the hit line 18 will remain precharged.

Figure 2 illustrates a first embodiment of a cache memory system 22. A content addressable memory 24 stores TAG values indicating the memory addresses of data values stored within corresponding cache memory rows 26 of the cache RAM block 28. Each content addressable storage row is formed of a plurality of content addressable memory cells 2. In Figure 2, a single row is schematically illustrated greatly vertically expanded.

35

5 A column of row drivers 30 are provided between the content addressable memory 24 and the cache RAM block 28. These are enabled at a time such that all the hit lines corresponding to non-matching rows will have discharged before any single remaining hit signal is passed to a row within the cache RAM block 28. A block of bit storage cells 32 storing valid and dirty bits associated with each row of
 10 the content addressable memory 24 and the cache RAM block 28 are also provided. The valid bit indicates that the row is storing valid data, as compared to at start-up when whatever values may be present within the system are not valid. The dirty bit is used in a writeback system to indicate that a cached value has been changed and needs to be written back to the main memory when it is flushed (removed) from the cache
 15 system.

Each row has an associated hit line 18 and access enable line 12. There is a further valid/dirty access enable line 34 for providing access to the valid/dirty bit block 32. A plurality of switchable buffers (e.g. AND gates) 36, 38, 40, 42 controlled
 20 by a multiplexer controller 44 serve to provide various interconnections between the hit line 18, the access enable line 12, the valid/dirty access enable line 34 and the output of an index decoder 46. The multiplexer controller 44 is responsive to an input mode signal Mode to select which of the gates 36, 38, 40 and 42 will be open and which will be closed in any given mode.

25

The index decoder 46 is responsive to an input index value to decode this and select one of the rows into which a new set of data values are to be written, e.g. to provide victim select.

30 In normal operation, when a read is to be made from the cache system 22, a portion of the virtual address value is applied to the content addressable memory 24 and compared in parallel to the TAG values stored within each content addressable storage row. If a match occurs, then a hit signal on the hit line 18 is generated for that row and is passed via the row driver block 30 to the cache RAM block 28. The cache
 35 row within the cache RAM block 28 is then read. During this operation all of the gates 36, 38, 40 and 42 may be closed.

5

When a write to a cache memory location takes place, then a similar comparison of the virtual address occurs within the content addressable memory 24 to generate a hit signal enabling write access within the cache row of the cache RAM block 28. In this case, the multiplexer controller 44 is responsive to the write mode to enable gate 38 to pass the hit signal so as to set the dirty bit for the row indicating that the value stored within that row has been changed and requires writing back to the main memory (not shown) in due course.

When a new value is being written to a content addressable storage row and a cache row, then the index decoder 46 selects the victim row using an index value from a victim generator circuit (not shown) e.g. a 6-bit random number. The gate 36 serves to pass the victim signal from the victim select circuit onto the hit line 18 to enable access to the corresponding cache data row. The gate 42 passes the victim select circuit to the access enable line 12 to enable the input data value upon the bit line pairs 14, 16 to be written to the content addressable storage row. The gate 40 is open to pass the victim select signal to set the valid bit for the row and reset the dirty bit.

In cache maintenance operations, such as cache clean and invalidate, there are two approaches that can be taken.

25

1. A portion of the virtual address is applied to the content addressable memory 24 and compared in parallel to the TAG values stored within each content addressable storage row. If a match occurs, then a hit signal on the hit line 18 is generated for that row and passed via gate 38 to the valid/dirty signal line 34. The valid and dirty bits are read, and if valid and dirty, then the signal on the hit line 18 is passed via the row driver block 30 to the cache RAM block 28. The cache data values for that row are then read. The valid and dirty bits are then written appropriately to reset the dirty bit and/or the valid bit, dependent upon whether a clean and/or invalidate operation is taking place.

35

5 2. The row can be selected by the index decoder 46. As before, the index select signal is passed onto the valid/dirty signal line 34 via gate 40 and the valid and dirty bits read. If valid and dirty, then the index select signal is passed via gate 36 to the hit line, and via the row driver block 30 to the cache RAM block 28. Cache data values for that row are then read. The valid and dirty bits are then written
10 appropriately for clean and/or invalidate operations.

In a diagnostic mode, such as a CAM read/write and/or a RAM read/write, then the system may operate in either of the modes described below.

15 1. The row is selected by the index decoder 46. The decoded index value is passed to the valid/dirty signal line 34 via gate 40 to the access enable signal line 12 via gate 42. The valid and dirty bits and the TAG held in the CAM for that row can then be read or written.

20 2. The row is selected by the index decoder 46. The decoded index value is passed on to the hit line 18 via gate 36 and via the row driver block 30 to the cache RAM block 28. The cache data word stored within the cache RAM block 28 can then be read or written.

25 Figure 3 illustrates a second embodiment of a cache memory system 22. This is a unified cache with writeback supported with the TAG value read out of the CAM 24 being used for writeback.

30 When it is desired to perform a block transfer of data then the following method may be used.

1. Linefill from a first address A (a LOAD from address A will be sufficient to initiate the linefill).
- 35 2. Perform a co-processor instruction to modify the TAG value for the row that has just been written to point to a second address B within the main memory

5 and mark the cache row as dirty in the valid/dirty block 32 to ensure that a writeback will subsequently occur.

3. The new data is now within the cache associated with the second address B and can be used directly. In the case of self-modifying code, the code can
10 be modified within the cache at the new address. The integrity of the main memory is ensured by the “dirty” data being written back to the external main memory when it is removed from the cache in due course.

The operation of the system in response to the co-processor instruction would
15 be first to find the matching row within the CAM 24 storing a TAG value matching the first address A. A hit signal for the matching row would then be asserted on the hit line 18 and latched by a latch 43. In a subsequent clock cycle the latch 43 would be enabled to drive its output onto the access enable signal line 12 to allow a new second address value B asserted upon the bit line pairs 14, 16 to be written to the
20 CAM cells 2 of the row concerned. The operation performed in response to the co-processor instruction would typically take approximately five clock cycles to complete.

This sequence of:

- 25 1. Linefill of 8 words from address A;
2. Co-processor instruction to change TAG from address A to address B;
is significantly shorter/faster than:
1. Load 8 words to the processor from address A (equivalent to a linefill of 8 words);
30 2. Store the 8 words to address B;
3. Linefill of 8 words from address B.

5

CLAIMS

1. Apparatus for processing data having a cache memory, said cache memory comprising:

10 a plurality of content addressable storage rows, each content addressable storage row containing a plurality of content addressable bit storage cells and serving to store a content addressable data word;

a plurality of bit lines connecting corresponding content addressable bit storage cells in different content addressable storage rows such that an input data word
15 on said plurality of bit lines is compared in parallel with the contents of each connected content addressable storage row;

a plurality of hit lines, each content addressable storage row having an associated hit line upon which a hit signal is asserted if said input data word matches a content addressable data word stored within said content addressable storage row;

20 a plurality of cache data rows each storing a row of cache data, said content addressable memory serving to store address data to identify memory addresses for corresponding rows of cache data such that a hit signal generated on a hit line in said content addressable memory enables access to a corresponding cache data row within said cache memory; and

25 an index decoder for decoding an index value to generate a select signal for selecting a content addressable storage row and a victim cache data row for replacement within said cache memory; wherein

said select signal is passed to said cache data row via a corresponding one of said hit lines.

30

2. Apparatus for processing data as claimed in claim 1, comprising a plurality of access enable lines, each content addressable storage row having an associated access enable line upon which an access signal may be asserted to allow a content addressable data word stored within said content addressable storage row to be
35 accessed.

5 3. Apparatus as claimed in any one of claims 1 and 2, wherein a valid bit is associated with each content addressable memory row for indicating whether said content addressable memory row is storing valid data.

10 4. Apparatus as claimed in claims 3, comprising a selectable connection for passing said hit signal to enable access to said valid bit.

15 5. Apparatus as claimed in any one of claims 1 and 2, wherein a dirty bit is associated with each content addressable memory row for indicating whether said cache data corresponding to said content addressable memory row has been changed since being written to said cache memory.

6. Apparatus as claimed in claims 5, comprising a selectable connection for passing said hit signal to enable access to said dirty bit.

20 7. A method of processing data using a cache memory, said method comprising the steps of:

storing content addressable data words within a plurality of content addressable storage rows, each content addressable storage row containing a plurality of content addressable bit storage cells;

25 connecting corresponding content addressable bit storage cells in different content addressable storage rows with a plurality of bit lines such that an input data word on said plurality of bit lines is compared in parallel with the contents of each connected content addressable storage row;

30 providing a plurality of hit lines, each content addressable storage row having an associated hit line upon which a hit signal is asserted if said input data word matches a content addressable data word stored within said content addressable storage row;

35 providing a plurality of cache data rows each storing a row of cache data, said content addressable memory serving to store address data identify memory addresses for corresponding rows of cache data such that a hit signal generated on a hit line in

5 said content addressable memory enables access to a corresponding cache data row within said cache memory; and

providing an index decoder for decoding an index value to generate a select signal for selecting a content addressable storage row and a cache data row for replacement within said cache memory; wherein

10 said select signal is passed to said cache data row via a corresponding one of said hit lines.

8. Apparatus for processing data having a cache memory, said cache memory comprising:

15 a plurality of content addressable storage rows, each content addressable storage row containing a plurality of content addressable bit storage cells and serving to store a content addressable data word;

a plurality of bit lines connecting corresponding content addressable bit storage cells in different content addressable storage rows such that an input data word on said plurality of bit lines is compared in parallel with the contents of each
20 connected content addressable storage row;

a plurality of hit lines, each content addressable storage row having an associated hit line upon which a hit signal is asserted if said input data word matches a content addressable data word stored within said content addressable storage row;

25 a plurality of access enable lines, each content addressable storage row having an associated access enable line upon which an access signal may be asserted to allow a content addressable data word stored within said content addressable storage row to be accessed;

a plurality of cache data rows each storing a row of cache data, said content
30 addressable memory serving to store address data identify memory addresses for corresponding rows of cache data such that a hit signal generated on a hit line in said content addressable memory enables access to a corresponding cache data row within said cache memory;

wherein, in an access mode, a connection is provided between a hit line for a
35 content addressable storage row and an access enable line for that content addressable storage row such that an input data word on said bit lines can be compared with said

5 content addressable data words within said plurality content addressable storage rows and if a match occurs for a content addressable storage row generating a hit signal to enable access to that row by passing said hit signal to serve as an access enable signal upon an access enable line.

10 9. Apparatus as claimed in claim 8, comprising an index decoder for decoding an index value to generate an access enable signal on one of said access enable lines.

10. Apparatus as claimed in claim 9, wherein said index value is used to select a victim for replacement within said content addressable memory.

15

11. Apparatus as claimed in any one of claims 8, 9 and 10, wherein a valid bit is associated with each content addressable memory row for indicating whether said content addressable memory row is storing valid data.

20 12. Apparatus as claimed in any one of claims 11, wherein said connection also passes said hit signal to enable access to said valid bit.

13. Apparatus as claimed in claim 8, wherein a dirty bit is associated with each content addressable memory row for indicating whether said cache data corresponding
25 to said content addressable memory row has been changed since being written to said cache memory.

14. Apparatus as claimed in any one of claims 13, wherein said connection also passes said hit signal to enable access to said dirty bit.

30

15. Apparatus as claimed in claims 13, wherein said data processing apparatus is responsive to a data processing instruction to use said access mode to:

identify a cache row storing cache data associated with first address data stored in a corresponding content addressable storage row;

35 store second address data in place of said first address data in said content addressable storage row; and

5 mark as dirty said cache data; such that

said cache data can be accessed from said cache memory using said second address data and said cache data is written back to a main memory location indicated by said second address data when said cache data is flushed from said main memory thereby effecting a transfer of said cache data from a main memory location indicated by said first address data to a main memory location indicated by said second address data.

16. A method of processing data using a content addressable memory, said method comprising the steps of:

15 storing content addressable data words within a plurality of content addressable storage rows, each content addressable storage row containing a plurality of content addressable bit storage cells;

connecting corresponding content addressable bit storage cells in different content addressable storage rows with a plurality of bit lines such that an input data word on said plurality of bit lines is compared in parallel with the contents of each connected content addressable storage row;

20 providing a plurality of hit lines, each content addressable storage row having an associated hit line upon which a hit signal is asserted if said input data word matches a content addressable data word stored within said content addressable storage row;

25 providing a plurality of access enable lines, each content addressable storage row having an associated access enable line upon which an access signal may be asserted to allow a content addressable data word stored within said content addressable storage row to be accessed;

30 wherein, in an access mode, a connection is provided between a hit line for a content addressable storage row and an access enable line for that content addressable storage row such that an input data word on said bit lines can be compared with said content addressable data words within said plurality content addressable storage rows and if a match occurs for a content addressable storage row generating a hit signal to enable access to that row by passing said hit signal to serve as an access enable signal upon an access enable line.

5

17. Apparatus for processing data substantially as hereinbefore described with reference to the accompanying drawings.

18. A method of processing data substantially as hereinbefore described with
10 reference to the accompanying drawings.



Application No: GB 9912630.2
Claims searched: 1-7

Examiner: Mike Davis
Date of search: 17 January 2000

Patents Act 1977
Search Report under Section 17

Databases searched:

UK Patent Office collections, including GB, EP, WO & US patent specifications, in:

UK Cl (Ed.R): G4A (AMA, AMC)

Int Cl (Ed.7): G11C, G06F

Other:

Documents considered to be relevant:

Category	Identity of document and relevant passage	Relevant to claims
	None	

X	Document indicating lack of novelty or inventive step	A	Document indicating technological background and/or state of the art.
Y	Document indicating lack of inventive step if combined with one or more other documents of same category.	P	Document published on or after the declared priority date but before the filing date of this invention.
&	Member of the same patent family	E	Patent document published on or after, but with priority date earlier than, the filing date of this application.



INVESTOR IN PEOPLE

Application No: GB 9912630.2
Claims searched: 8-16

Examiner: Mike Davis
Date of search: 3 May 2000

Patents Act 1977
Further Search Report under Section 17

Databases searched:

UK Patent Office collections, including GB, EP, WO & US patent specifications, in:

UK Cl (Ed.R): G4A (AMA, AMC)

Int Cl (Ed.7): G11C, G06F

Other:

Documents considered to be relevant:

Category	Identity of document and relevant passage	Relevant to claims
	None	

X	Document indicating lack of novelty or inventive step	A	Document indicating technological background and/or state of the art.
Y	Document indicating lack of inventive step if combined with one or more other documents of same category.	P	Document published on or after the declared priority date but before the filing date of this invention.
&	Member of the same patent family	E	Patent document published on or after, but with priority date earlier than, the filing date of this application.